

Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure

L. ORSINI,* J. CORANDER,† A. ALASENTIE* and I. HANSKI*

*Metapopulation Research Group, Department of Biological and Environmental Sciences, University of Helsinki, PO Box 65 (Viikinkaari 1), FIN-00014 Helsinki, Finland, †Department of Mathematics, Åbo Akademi University, FIN-20500 Åbo, Finland

Abstract

The Glanville fritillary butterfly (*Melitaea cinxia*) has been studied in the Åland Islands in Finland since 1991, where it occurs as a classic metapopulation in a large network of 4000 dry meadows. Much ecological work has been conducted on this species, but population genetic studies have been hampered by paucity of suitable genetic markers. Here, using single nucleotide polymorphisms and microsatellites developed for the Glanville fritillary, we examine the correspondence between the demographic and genetic spatial structures. Given the dynamic nature of the metapopulation, the current genetic spatial structure may bear a signal of past changes in population sizes and past patterns of gene flow rather than reflect the current demographic structure or landscape structure. We analyse this question with demographic data for 10 years, using the Rand index to assess the similarity between the genetic, demographic, and landscape spatial structures. Our results show that the current genetic spatial structure is better explained by the past rather than by the current demographic spatial structure or by the spatial configuration of the habitat in the landscape. Furthermore, current genetic diversity is significantly explained by past metapopulation sizes. The time lag between major demographic events and change in the genetic spatial structure and diversity has implications for the study of spatial dynamics.

Keywords: demographic perturbation, fragmented landscape, gene flow, Glanville fritillary, microsatellite, null allele, SNP, spatial population structure

Received 2 November 2007; revision received 29 January 2008; accepted 11 March 2008

Introduction

Many species occur as metapopulations in fragmented landscapes, consisting of discrete local populations connected by migration, the magnitude of which depends on landscape structure and habitat quality as well as on the migration propensity of the species (Hanski 1999; Hanski & Gaggiotti 2004). The long-term persistence of such metapopulations hinges on a stochastic balance between local extinctions and recolonizations (Levins 1969; Hanski 1999). Parallel to the discontinuous demographic population structure, the genetic composition of metapopulations is

likely to exhibit spatial structure, as limited gene flow may allow local populations (often called demes) and clusters of such populations to differentiate due to founder effects and drift. The genetic composition of metapopulations may additionally be affected by spatially varying selection pressures. It is of interest to ask how the spatial structure of habitat in a fragmented landscape, the demographic structure of the metapopulation inhabiting that landscape, and the genetic spatial structure of the metapopulation are related to each other.

Studies of spatial genetic structures are appealing because patterns can be described at any spatial scale from individuals to continents (Manel *et al.* 2003), and these patterns can reveal the influence of processes operating on temporal scales from the present to millennia (Slatkin 1985; Davies *et al.* 1999). Nonetheless, interpretations can be challenging, because different population processes can lead to similar

Correspondence: Luisa Orsini, Metapopulation Research Group, Department of Biological and Environmental Sciences, PO Box 65 (Viikinkaari 1) 00014 University of Helsinki, Finland. Fax: +358 9191 57694; E-mail: address: luisa.orsini@helsinki.fi

genetic patterns (Slatkin 1985; Ray 2001). Landscape structure plays an important role in determining spatial genetic structures, but this influence is mediated via population processes and the patterns represent the cumulative outcome of gene flow in the past (Burel *et al.* 1998). The demographic and genetic spatial population structures may be congruent, especially when populations are isolated and stable, but in metapopulations local populations are often ephemeral and there is a spatially variable amount of migration and gene flow among populations. Local populations may have largely independent demographic dynamics, but often the dynamics are spatially correlated, typically reflecting some large-scale environmental effects (Ovaskainen & Hanski 2004; Hanski & Meyke 2005). In this situation, an integrated approach combining knowledge of the ecological processes with inference from genetic markers can provide a more robust insight into population processes operating across landscapes than studies relying on one type of information only.

These questions also raise the issue of how the genetic spatial structures should be quantified when dealing with species living in fragmented landscapes. Many authors have studied the effect of habitat fragmentation (Coulon *et al.* 2004 and references therein) using Wright's *F*-statistics (Wright 1931) or assignment tests among well-defined populations. Unfortunately, for complex genetic structures in populations extending across heterogeneous landscapes, it is difficult to extract useful information using these approaches, and it is challenging to relate such results to additional data on relevant demographic processes, because of spatially varying connectivity, and hence, gene flow among populations. The recently developed Bayesian spatial clustering methods (Guillot *et al.* 2005; Corander *et al.* 2008) provide a way of inferring statistically a spatially explicit representation of the genetic structure, which can be efficiently compared with landscape and demographic structures.

Here, we analyse the spatial genetic structure in the Glanville fritillary butterfly (*Melitaea cinxia*). This butterfly occurs as a classic metapopulation in a highly fragmented landscape in the Åland Islands in Southwest Finland. Much ecological work has been conducted on this species, including empirical (reviewed by Hanski 1999; Ehrlich & Hanski 2004) and modelling studies on spatial structures and dynamics (Hanski 1994; Hanski *et al.* 1996; Ovaskainen & Hanski 2003; Ovaskainen & Hanski 2004), but corresponding population genetic studies (Saccheri *et al.* 2004) have been hampered by paucity of suitable genetic markers. Development of microsatellite markers has been difficult for this species (Palo *et al.* 1995; Sarhan 2006) as for other Lepidoptera (Keyghobadi *et al.* 1999; Neve & Meglecz 2000; Keyghobadi *et al.* 2005a, b). Recently, we have developed single nucleotide polymorphisms (SNP) for the Glanville fritillary (Orsini *et al.* 2007) and conducted a large-scale genetic study of the metapopulation. Our aim in this study

is to examine the correspondence between the spatial configuration of the habitat and the demographic and genetic spatial structures. Given the highly dynamic nature of metapopulations (Hanski 1999), the current genetic spatial structure may bear a signal of past changes in population sizes and past patterns of gene flow rather than reflect the current spatial demographic structure or the spatial configuration of the habitat in the landscape. We analyse this question with demographic data for 10 years sampled across several hundred local populations. This study contributes towards a more mechanistic understanding of the effects of demographic dynamics in fragmented landscapes on genetic population structure.

Materials and methods

The Glanville fritillary metapopulation and sampling

The long-term study of the Glanville fritillary butterfly in the Åland Islands in Southwest Finland was started in 1991 (Hanski 1999). In 1993, and again in 1998–1999, the entire study area was mapped for suitable habitat for the butterfly. The habitat consists of small dry meadows (average area 0.15 ha) with one or both of the larval host plants, *Plantago lanceolata* and *Veronica spicata* (Hanski 1999; Nieminen *et al.* 2004). The number of suitable habitat patches is *ca* 4000.

Females oviposit in large clusters of 150–200 eggs (Saastamoinen 2007), and the larvae stay most of their development in sib-groups. Since 1993, all habitat patches have been monitored for the presence and the sizes of local populations based on a count of the larval groups in late summer (Nieminen *et al.* 2004). Every year, 500–700 of the 4000 meadows have been inhabited by the butterfly, as evidenced by the presence of larvae (Hanski 1999; Nieminen *et al.* 2004). The rate of population turnover is high, with about 100 local populations going extinct in each year and roughly the same number of new populations being established by dispersing butterflies. The populations that turn over are mostly the smallest ones, but the network has no large habitat patches and hence, all local populations have a significant risk of extinction (Hanski *et al.* 1995; Hanski 1999). Very few populations have persisted continuously since 1993.

For the present study, we sampled all local populations in the northern part of the Åland Islands in the autumn 2002, covering about half of the entire range of the Glanville fritillary. One larva was sampled from each larval group that was detected, yielding a total sample of 737 larvae, sampled from 186 local populations. Previous studies indicate that about 50% of all existing larval groups are detected in the survey (Nieminen *et al.* 2004), hence, our sample includes about half of the larval groups in the metapopulation at the time of sampling. Samples were preserved in alcohol at –20 °C before DNA extraction.

| Locus | Primer 5'–3' | SNP type | PCR size (bp) | T_m °C | Multiplex |
|-------|------------------------|----------|---------------|----------|-----------|
| 30.1 | ATAAAATTAATTACATACTG | A/C | 298 | 59 | 1 |
| 42.1 | AGTCAATTTTGTACCCATT | A/G | 184–187 | 59 | 1 |
| 53.2 | ATAAAATTCACATTTATATG | A/G | 245–256 | 59 | 1 |
| 3 N | TTATCAGTTTGTGCATTAAT | T/A | 305–319 | 65 | 2 |
| C60 | ACTTTAACTGTGATAAATC | A/G | 328–331 | 60 | 3 |
| C98 | ACCATAATTGTGTATATATCT | A/T | 298–315 | 60 | 3 |
| C113 | ACCCCTCTCTCTCCCGTGG | C/T | 324–328 | 60 | 4 |
| C120 | ATCCGTTCTTATTAAACCC | C/T | 286–294 | 60 | 4 |
| C129 | TAAGCACTCTTCTGTTTT | T/G | 204 | 60 | 4 |
| C152 | TAAAACAGTTTAACGAGTTCTG | T/C | 310 | 65 | 2 |

Table 1 Primers used for SNP genotype assay by primer extension. The table shows the sequence of the screening primer, PCR fragment size and the optimal annealing temperature (T_m). Loci that were coamplified are indicated as multiplex groups. The markers are described in Orsini *et al.* (2007)

In the present study, we use data only for the populations that occurred on meadows that have been monitored continuously since 1993, omitting populations in habitat patches that have been discovered since 1993. Thus the results are comparable for the 10-year period 1993–2002.

DNA isolation and microsatellite and SNP typing

Genomic DNA was isolated using Nucleo spin tissue extraction kit (Mackerey-Nagel), with overnight incubation at 56 °C. All samples were genotyped for five microsatellites and 10 SNPs. The microsatellites have been previously isolated from the Glanville fritillary (Sarhan 2006). SNPs have been isolated using a rapid and cost-effective strategy exploiting the property of the bacteriophage Mu transposition machinery to target mismatched DNA sites (Orsini *et al.* 2007). We used 10 of the 24 SNPs that have been discovered, choosing the ones located on independent genomic regions and hence, unlinked (Table 1).

Microsatellite loci were amplified in two multiplex polymerase chain reactions (PCR) that combined the loci CINXIA1A–CINXIA82 (PCR 1) and CINXIA19–CINXIA35–CINXIA77 (PCR 2), respectively. Sequences for several individuals demonstrated that CINXIA82 amplified a doublet fragment, representing two independent loci, and therefore this marker was not used further.

Ten to 20 ng of genomic DNA were amplified in PCRs consisting of 10× NH₄SO₄ PCR buffer, 0.5/1 µM of each primer, 200 µM of each of the dNTPs, 2.5 mM of MgCl₂, 20 ng of bovine serum albumin (BSA) and 0.1 U of *Taq* Polymerase (Fermentas, Life Sciences). The primer concentration was adjusted according to the amplification efficiency in the presence of multiple primer pairs. All amplifications were performed in 10-µL final volumes using MJ DNA Engine Tetrad 2 (MJ Research, Peltier). PCR cycling conditions were as follows: denaturation at 94 °C for 3 min followed by 30 cycles of 94 °C for 1 min, 56 °C (PCR 1) or 59 °C (PCR 2) for 1 min and 72 °C for 1 min. One primer of each primer set (typically the reverse primer) was end-labelled with a fluorescent dye, either NED (CINXIA19), FAM (CINXIA77

and CINXIA1A), or HEX (CINXIA35). The alleles were scored using MegaBACE 1000 automated sequencer (GE Healthcare) and analysed with MEGABACE FRAGMENT PROFILER software (GE Healthcare).

SNPs were genotyped by primer extension reactions (Solokov 1990), in which the screening primers, designed with a 3' end immediately adjacent to the SNP, undergo a single nucleotide extension by a fluorescent-labelled ddNTP corresponding to the SNP allele. To reduce the number of reactions required, the 10 SNPs were amplified in four multiplex PCRs, in which loci were combined according to their optimal annealing temperature and PCR length size (Table 1). Each PCR (20 µL) contained 20–30 ng of genomic DNA, 1 µM each primer, 200 µM each dNTPs, 2.5 mM MgCl₂, 20 ng BSA, and 0.2 U *Taq* DNA Polymerase (Fermentas, Life Sciences). An initial denaturing step (5 min at 95 °C) was followed by 35 cycles of amplification with 1 min at 94 °C, 1 min at the annealing temperature, and 1.5 min at 72 °C. A final extension step included incubation for 15 min at 72 °C. PCR products were purified with Exo-SAP-IT (GE Healthcare) at the concentration of 1 µL/10 µL PCR. Primer extension reactions employed the SNUPe kit (GE Healthcare, Batley, Hayes, 2003) following the manufacturer's instructions, and were run on MegaBACE 1000 (GE Healthcare).

Genotypes were called by SNP PROFILER (GE Healthcare) and the calls were checked visually. Screening primer sequences are shown in Table 1. For both microsatellite and SNP loci, the samples that failed were repeated together with random successful samples. Samples that failed twice were considered as missing data. To reduce human error in reading genotyping calls, the calls were scored independently by two people and the calls were compared.

Statistical analyses

For each microsatellite locus, the expected and observed heterozygosities were calculated and genotype frequencies were tested against the Hardy–Weinberg expectation using ARLEQUIN (Excoffier *et al.* 2005). Because we found excess homozygosity, we searched for evidence of null alleles using

the program MICRO-CHECKER (van Oosterhout *et al.* 2004). Evidence for null alleles was found in three of the four microsatellites (CINXIA19, CINXIA77 and CINXIA1A). The program FREENA (Chapuis & Estoup 2007) was used to calculate the frequency of null alleles at different loci and F_{ST} values were re-calculated after correcting for the presence of null alleles. F_{ST} were calculated with the Weir and Cockerham (Weir & Cockerham 1984) method for pairs of populations using MSA (Dieringer & Schlötterer 2003) and 1000 nonparametric bootstrap replicates.

For the SNPs, the observed and expected heterozygosities and the F_{ST} values were calculated using ARLEQUIN (Excoffier *et al.* 2005). An exact test was calculated for the departure from the Hardy–Weinberg equilibrium and the significance was determined by the Markov Chain Monte Carlo (MCMC) simulation available in ARLEQUIN (Guo & Thompson 1992).

For both marker types, expected and observed heterozygosities were calculated for the pooled sample (186 populations) as well as for a subsample omitting the smallest populations, which are prone to rapid drift (< 4 larval groups). For both marker types, the significance of correlations between genetic and geographical distances was calculated using the Mantel test (Mantel 1967) with 10 000 matrix randomizations (GENEPOP, Raymond & Rousset 1995). Isolation by distance was investigated using the pooled sample as well as the subsample omitting the smallest populations (< 4 larval groups). We repeated the calculations after removing one microsatellite locus at a time to check that the results were not due to any particular locus.

Genetic and demographic clusterings

The spatial genetic structure was described using the Bayesian analysis implemented in the BAPS software (Guillot *et al.* 2005; Corander & Marttinen 2006; Corander *et al.* 2008). BAPS can fit several Bayesian models of population structure to molecular data. The clusters thus obtained represent genetically differentiated parts of the entire metapopulation. We employed the spatial model option in BAPS, using local populations inhabiting discrete habitat patches with known geographical coordinates as the population units to be clustered. All molecular data collected from a particular local population were used to obtain the posterior distribution of allele frequencies for that population.

Under the spatial model, the genetic structure is calculated assuming *a priori* that the structure within a particular area depends on the neighbouring areas. Corander *et al.* (2008) have shown that the spatial model improves the statistical power to detect the underlying population structure when the molecular data are sparse. In the spatial clustering model, the landscape occupied by the population is divided into a 'coloured Voronoi tessellation' (Deussen *et al.* 2000), in which different colours in the tessellation represent genetically differentiated populations. Identical colour

for particular tessellation cells is inferred through Bayes' formula using both molecular data and a spatial prior distribution, which expresses expected dependence in the colours of neighbouring cells. The model specifies the colouring, corresponding to an estimate of the underlying genetic population structure, jointly for all cells of the tessellation. A stochastic search algorithm is used to estimate the maximum *a posteriori* colouring. In the present application, the spatial model was fitted using 20 replicate runs, each with the maximum number of 25 clusters. This *a priori* upper bound was judged to cover the range of clusters that could be statistically detected in this data set. The resulting maximum *a posteriori* estimate for the number of genetic clusters (see Results) was well below this threshold, and hence the value of 25 clearly exceeded the number of clusters detected in these data. We pooled the two sets of markers to maximize information for genetic clustering.

Demographic clustering based on population sizes and their spatial locations was calculated with the average linkage hierarchical clustering algorithm (Seber 1984). The algorithm was applied to a distance matrix, in which the distance between local populations *i* and *j* was calculated as the average of their reciprocal pairwise connectivity values, defined for patch *i* as (Hanski 1999)

$$S_i = N_j e^{-d_{ij}},$$

where N_j is population size (number of larval groups) in patch *j* and d_{ij} is the Euclidian distance between patches *i* and *j* in kilometres. This formula assumes that average migration distance is 1 km, which is supported by empirical data (Hanski 1999). Essentially, S_i measures the expected flow of migrants from population *j* to population *i*. Connectivity values were calculated separately for each year because population sizes vary from one year to another.

We examined year-to-year similarity of the demographic clusterings using the adjusted Rand index (RI; Hubert & Arabie 1985). RI is a probabilistic measure of the degree of similarity between two clusterings of the same set of objects. When RI equals zero, the similarity between the two clusterings is at the level one would expect by a comparison of random partitions overlapping by chance. The higher the RI value, the higher the probability that a randomly chosen pair of objects is concordantly clustered in the two clusterings. The Rand index has been shown to be conservative, which means that it is not easily inflated towards unity unlike many other related measures of clustering concordance. In our analysis, large values of RI indicate little change in the overall spatial configuration of populations.

The demographic clustering was obtained using a hierarchical clustering algorithm, as described above, which produces a dendrogram rather than a partition of the habitat patches. The dendrogram was cut at the required level to obtain a certain number of clusters. These clusters are

groupings of habitat patches based on their geographical location in the landscape and their current population sizes. The 'required level' refers to the cut-off level in the dendrogram that would yield the desired number of clusters. For each year, we produced demographic clusterings with 10–40 clusters, calculated the year-to-year comparison for each one of them, and used the average of the resulting RI values as the measure of demographic similarity between the two years.

We calculated similarity between the genetic clustering for 2002 and demographic clusterings for different years using the RI. This was carried out in two different ways, by comparing the genetic clustering with either the same number of demographic clusters or with the 'optimal' number of demographic clusters, optimal meaning the number of demographic clusters that produced the highest RI value with the genetic clustering for 2002. Finally, we compared the genetic clustering with the spatial structure of the habitat, which was calculated with the average linkage hierarchical clustering algorithm as above for the demographic clustering, but now the distance matrix consisted of just the pairwise Euclidean distances between habitat patches, ignoring any information on the presence of butterfly populations. Thus, connectivity of patch i was now calculated as

$$S_i = e^{-d_{ij}}.$$

The hierarchical distribution of genetic variation in the metapopulation was described with a three-level analysis of molecular variance (AMOVA) using ARLEQUIN (Excoffier *et al.* 2005). The analysis was repeated for the combination of microsatellites and SNPs as well as for each marker type separately. The three levels were the genetic clusters as identified by BAPS, populations within clusters, and individuals (representing larval groups) within populations. Within-individual variation was estimated by permuting all individual genotypes in the entire metapopulation. Permutation tests (10 000 permutations) were used to assess whether genetic differentiation at the different levels was significantly greater than zero (Excoffier *et al.* 2005).

Results

Demographic metapopulation structure

The total number of distinct local populations in the study area varied from 97 to 202 in the years from 1993 to 2002, taking into account only those habitat patches that have been monitored continuously since 1993 (Material and methods). The corresponding range in the pooled number of larval families was from 383 to 1149. The size of the metapopulation has remained relatively stable with a slightly increasing trend (Fig. 1a). The greatest increases in

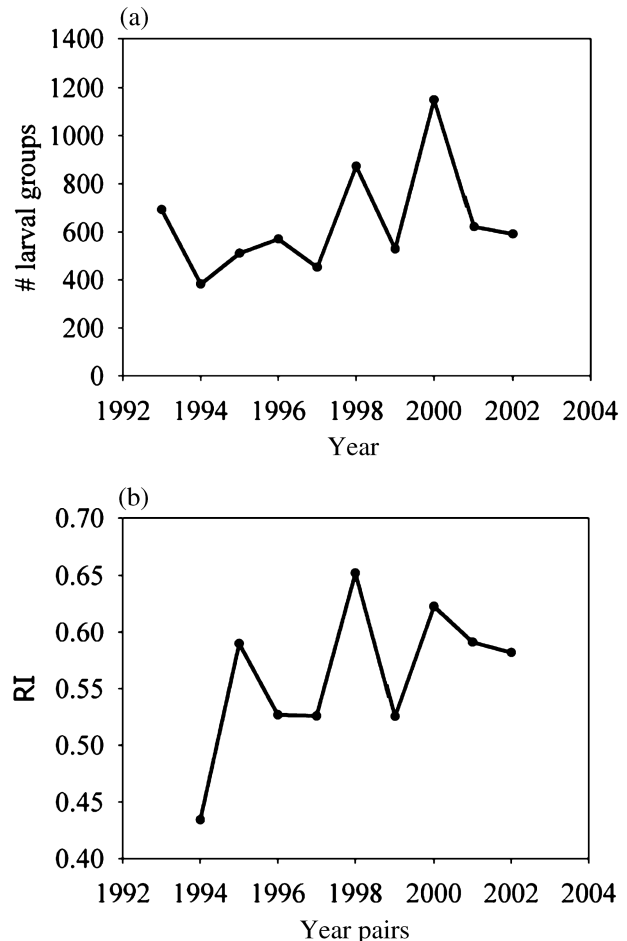


Fig. 1 (a) Change in metapopulation size as measured by the pooled number of larval groups in all local populations from 1993 to 2002. (b) Similarity of the demographic structures between successive years as measured by the Rand index (Materials and methods).

metapopulation size occurred at the two transitions in 1997–1998 and 1999–2000 (Fig. 1a; see also Fig. 4.12 in Nieminen *et al.* 2004).

Figure 2 shows the yearly changes in the sizes and spatial distribution of local populations from 1993 to 2002. The demographic dynamics have been markedly spatially correlated, as described in detail by Nieminen *et al.* (2004) and Hanski & Meyke (2005), although patterns have been dissimilar in different transitions. The most important reason for the spatially correlated changes in population sizes is spatially correlated weather effects (Hanski & Meyke 2005). The pattern in the occurrence of local populations became spatially less even and shifted towards northeast in 1997–1998 (Fig. 2; see also Fig. 1 in Hanski & Meyke 2005).

We examined year-to-year similarity in the demographic structure using the RI, based on the spatial clustering of populations. RI was high in 1997–1998 and from 1999

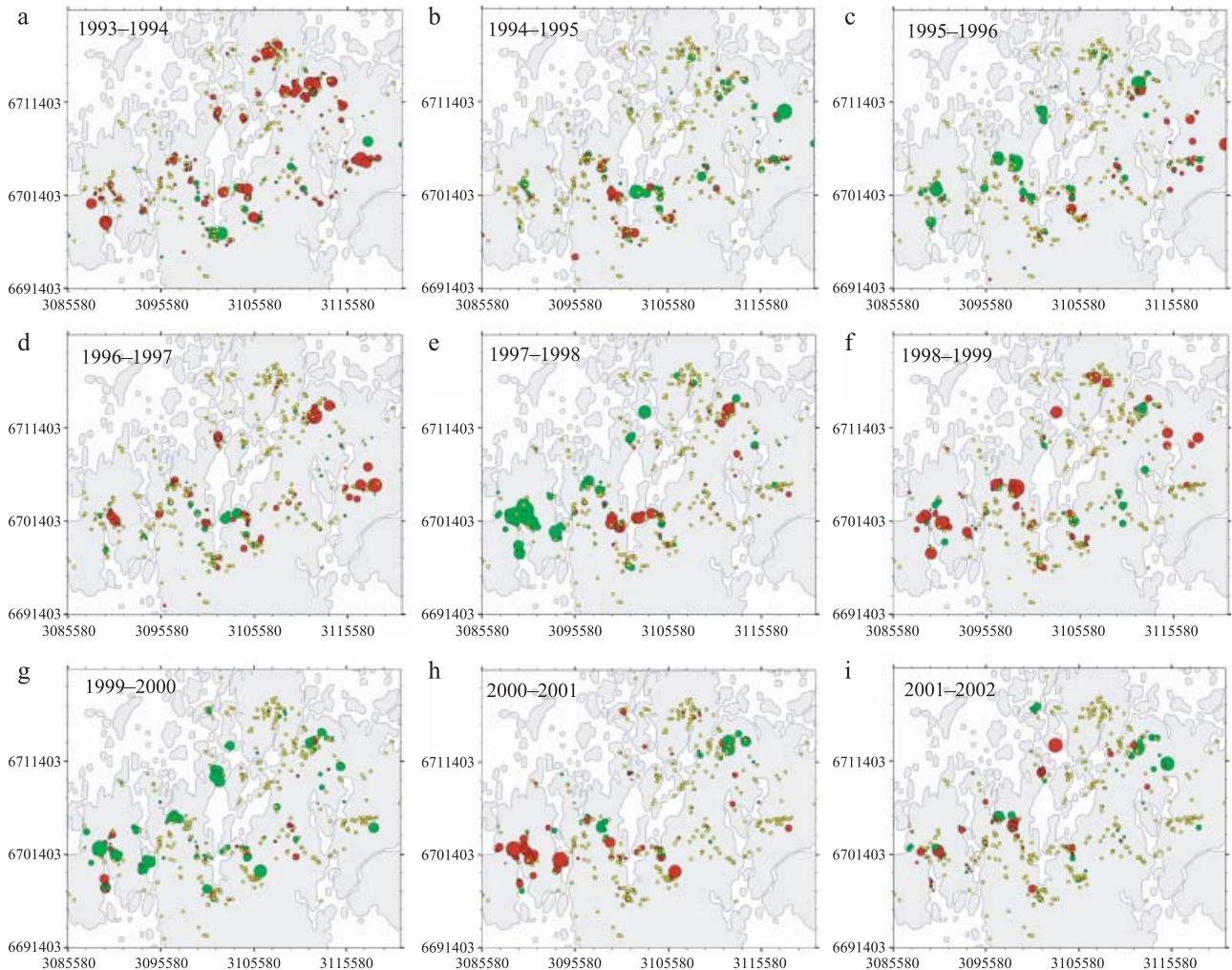


Fig. 2 (a–i) Yearly transitions in the sizes of local populations. The green colour indicates an increase and the red colour a decrease in population size on a logarithmic scale. The yellow circles represent populations with unchanged size.

onwards, while it was low in 1998–1999 and in the early transitions (Fig. 1b). There is no simple relationship between the RI values (Fig. 1b) and changes in the pooled size of the metapopulation (Fig. 1a), but it is noteworthy that high similarity tended to coincide with an increase in overall metapopulation size. Conversely, low similarity in the demographic structures at the transition 1998–1999 coincides with a large drop in metapopulation size. These results probably reflect dissimilar patterns of population turnover with increasing and decreasing size of the entire metapopulation.

Polymorphism and linkage disequilibrium of genetic markers

The level of polymorphism in the microsatellites ranged from 2 to 32 alleles per locus, with an average of 15.3. An

overall deficit of heterozygosity was evident in the pooled sample as well as in all four loci when analysed separately (Table 2). Comparison of the observed genotypes with the distribution of randomized genotypes generated with the program MICRO-CHECKER (van Oosterhout *et al.* 2004) suggested that heterozygote deficiency is due to the presence of null alleles, which are not uncommon in butterflies (Palo *et al.* 1995; Keyghobadi *et al.* 1999; Meglecz *et al.* 2004; Sarhan 2006).

Six of the 10 SNPs showed a significant deviation from the Hardy–Weinberg equilibrium (Table 2). In one case, the observed heterozygosity was greater than expected (locus 42.1), whereas in the remaining five SNPs, the observed heterozygosity was lower than expected (Table 2). Once again, the likely reason for the latter result is the presence of null alleles, probably due to variation at the primer sites. The results for both microsatellites and SNPs were broadly

| | SNPs | | | | Microsatellites | | | |
|-----------------------------------|-------|-------|-------|-----------------|-----------------|-------|-------|-----------------|
| | Locus | H_O | H_E | P value | Locus | H_O | H_E | P value |
| All populations ($n = 186$) | 30.1 | 0.31 | 0.32 | 0.640 | CINXIA 1 A | 0.34 | 0.89 | < 0.0001 |
| | 42.1 | 0.64 | 0.50 | < 0.0001 | CINXIA 19 | 0.43 | 0.75 | < 0.005 |
| | 53.2 | 0.13 | 0.15 | 0.002 | CINXIA 35 | 0.36 | 0.42 | < 0.0001 |
| | 3 N | 0.28 | 0.29 | 0.097 | CINXIA 77 | 0.59 | 0.82 | < 0.0001 |
| | C60 | 0.27 | 0.42 | < 0.0001 | | | | |
| | C98 | 0.42 | 0.48 | 0.00069 | | | | |
| | C113 | 0.38 | 0.37 | 0.84 | | | | |
| | C120 | 0.45 | 0.47 | 0.185 | | | | |
| | C129 | 0.14 | 0.47 | < 0.0001 | | | | |
| | C152 | 0.33 | 0.50 | < 0.0001 | | | | |
| Large populations ($n = 80$) | 30.1 | 0.30 | 0.31 | 0.227 | CINXIA 1 A | 0.34 | 0.89 | < 0.0001 |
| | 42.1 | 0.65 | 0.49 | < 0.0001 | CINXIA 19 | 0.42 | 0.76 | < 0.0001 |
| | 53.2 | 0.14 | 0.15 | 0.093 | CINXIA 35 | 0.37 | 0.44 | < 0.0001 |
| | 3 N | 0.27 | 0.30 | 0.071 | CINXIA 77 | 0.59 | 0.82 | < 0.0001 |
| | C60 | 0.29 | 0.42 | < 0.0001 | | | | |
| | C98 | 0.43 | 0.48 | 0.014 | | | | |
| | C113 | 0.36 | 0.36 | 1.00 | | | | |
| | C120 | 0.44 | 0.47 | 0.104 | | | | |
| | C129 | 0.15 | 0.46 | < 0.0001 | | | | |
| | C152 | 0.34 | 0.50 | < 0.0001 | | | | |

Table 2 Population genetic statistics for microsatellite and SNP loci. H_O and H_E are the observed and expected heterozygosities, while the P value gives the significance of the deviation from the Hardy-Weinberg equilibrium. Data are shown for all populations and for populations larger than three larval groups. Bold face indicates significant departure from the Hardy-Weinberg equilibrium

the same when the smallest local populations with < 4 or < 10 larval groups were excluded (Table 2 shows the results when local populations with < 4 larval groups were excluded).

Deviations from the Hardy-Weinberg equilibrium could be due to spatial heterogeneity in allele frequencies. We tested this possibility by repeating the calculations separately for the genetic clusters (which are described below) with the largest pooled sample of individuals (details in Table S1, Supplementary material). Of the four microsatellites, three showed significant heterozygote deficit in all six clusters, but CINXIA35 showed a significant deficit in only one genetic cluster. Of the 10 SNPs, five showed significant deviations in at least two genetic clusters (Table S1). These SNPs were largely the same that showed significant deviations also in the pooled data (Table 2), but intriguingly, SNPs 30.1 and 3N showed a significant deviation in two and three clusters, respectively, although not in the pooled data. SNPs 53.2, C113, and C120 showed no deviations from the Hardy-Weinberg equilibrium in any of the genetic clusters.

There was no significant linkage disequilibrium, with the exception of the microsatellite locus CINXIA35 exhibiting linkage disequilibrium with CINXIA1A and CINXIA19.

Genetic spatial structure

We used the pooled data for microsatellites and SNPs to infer the genetic clustering of local populations, using infor-

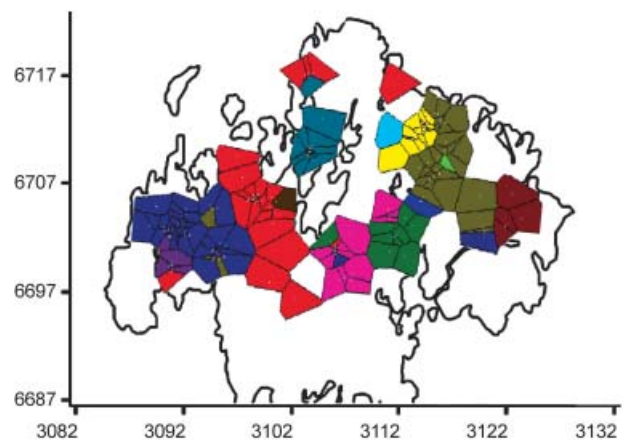


Fig. 3 Genetic clustering of local populations in the study area in the Åland Islands inferred with the program BAPS and using SNP and microsatellite markers. Identical colours identify populations with homogeneous genetic composition, whereas different colours represent genetically differentiated populations. There are altogether 13 homogeneous clusters.

mation on the spatial locations of the populations as a prior (Material and methods). The highest marginal likelihood (corresponding to maximizing the posterior probability) was obtained for 13 clusters (Fig. 3). The stability of this solution was investigated by running the stochastic estimation many times and observing where the algorithm converged. The solution with $k = 13$ was obtained in most cases. The solutions can be compared analytically with BAPS,

Table 3 Three-level hierarchical AMOVA (genetic clusters, populations, larval groups) for pooled microsatellites and SNPs, and for the two markers separately, based on estimates of genetic distances as calculated in ARLEQUIN (Excoffier *et al.* 2005). Significant differentiation ($P < 0.001$) based on permutation tests (10 000 permutations) is marked with an asterisk (*)

| Source of variation | d.f. | Microsatellites and SNPs | | | Microsatellites | | SNPs | |
|---------------------|------|--------------------------|-------------------------|-------------------|-------------------------|-------------------|-------------------------|-------------------|
| | | Variance component | Percentage of variation | Fixation indices | Percentage of variation | Fixation indices | Percentage of variation | Fixation indices |
| Among clusters | 12 | 0.12 Va | 4.94 | $F_{ct} = 0.05^*$ | 5.56 | $F_{ct} = 0.06^*$ | 4.53 | $F_{ct} = 0.04^*$ |
| Among populations | 173 | 0.06 Vb | 2.58 | $F_{sc} = 0.03^*$ | 1.22 | $F_{sc} = 0.01^*$ | 3.47 | $F_{sc} = 0.04^*$ |
| Among larval groups | 550 | 0.14 Vc | 6.02 | $F_{ls} = 0.06^*$ | 22.18 | $F_{ls} = 0.24^*$ | -4.61 | $F_{ls} = -0.05$ |
| Within individual | 736 | 2.05 Vd | 86.47 | $F_{it} = 0.13^*$ | 71.04 | $F_{it} = 0.29^*$ | 96.91 | $F_{it} = 0.03$ |

and in these comparisons the mode estimate was found to be very stable. Examining the posterior neighbourhood of the estimate further confirmed the overall statistical stability of the estimated clustering solution. The clusters are mostly discrete and the larger ones are 5–10 km across (Fig. 3). Clusterings inferred from microsatellites and SNPs separately gave congruent results but with reduced resolution when SNPs alone were used (results not shown).

A three-level hierarchical AMOVA for the combination of microsatellites and SNPs showed high levels of genetic differentiation among larval groups and among genetic clusters (Table 3). Permutation tests yielded significant results at all levels (Table 3). Results were similar for the two marker types separately, with the exception of no within-population among-larval group genetic differentiation for SNPs (Table 3). The presence of only two alleles in SNPs reduces the power to detect structure at the within-population level.

Average F_{ST} value over all loci was 0.02, with the locus CIXIA19 being the most divergent marker (0.04). The Mantel test did not detect significant isolation by distance for the two markers combined ($P = 0.98$) nor when they were analysed separately ($P = 0.26$ for microsatellites and $P = 0.43$ for SNPs). There was a marginally ($r = 0.008$; $P < 0.05$) significant relationship when only populations larger than three larval groups were analysed (data not shown). Drift in small populations may explain why the exclusion of the smallest populations gives a stronger signal of isolation by distance.

Comparisons between landscape structure and demographic and genetic spatial structures

Similarity between the genetic and demographic spatial structures (clusterings) was assessed with RI in two different ways. First, the genetic clustering for 2002 (13 clusters, Fig. 3) was compared with the 'optimal' demographic clustering for each year, meaning the number of demographic clusters that produced the highest RI value with the genetic structure for 2002. The number of such demographic clusters varied

from 7 to 24. Second, the clusterings were compared when the number of demographic clusters was forced to be the same as the number of genetic clusters. In this case, the number of genetic clusters ranged from 6 to 13 in different years, because in all years, except in 2002, some of the 2002 genetic clusters had no corresponding populations (Fig. 4).

The results show substantial variation from year-to-year (dotted lines, Fig. 4a, b) but also a pattern that becomes clearer in a 3-point moving average of the yearly values (continuous lines, Fig. 4a, b). The match between the demographic and genetic clusterings increases backwards in time down to 1995 or 1996, beyond which it again decreases (Fig. 4). To assess whether this pattern could be due to chance (drift) only, we constructed the following test for the pattern in Fig. 4(b). First, we fitted a Gaussian kernel function to predict the moving average smoothed RI values as a function of the distance (in years) from the year in which the maximum smoothed RI was obtained. The fitted function yields the correct smoothed RI value in the year with the maximum value, and the predicted RI value decreases exponentially backwards and forwards in time, such that the rate of decrease is determined by the least squares minimization of the residual sum of squares (RSS) with respect to the observed (smoothed) RI values. This yields an observed value, say T , for the discrepancy between the fitted function and the data. Next, we fitted an analogous function to moving average smoothed values calculated for 10 000 permutations of the yearly RI values. Finally, a P value was obtained by calculating the probability $P(RSS > T)$ based on the 10 000 sets of RI values. The P value was 0.011, supporting the conclusion that the unimodal pattern in Fig. 4 is not due to chance.

The genetic structure was compared with the spatial configuration of the habitat in the landscape, not including any information about the populations (Materials and methods; dashed lines, Fig. 4). The correspondence between these clusterings is invariant in time and is higher than the match between the demographic and genetic clusterings for 2002, but lower than the match between the genetic

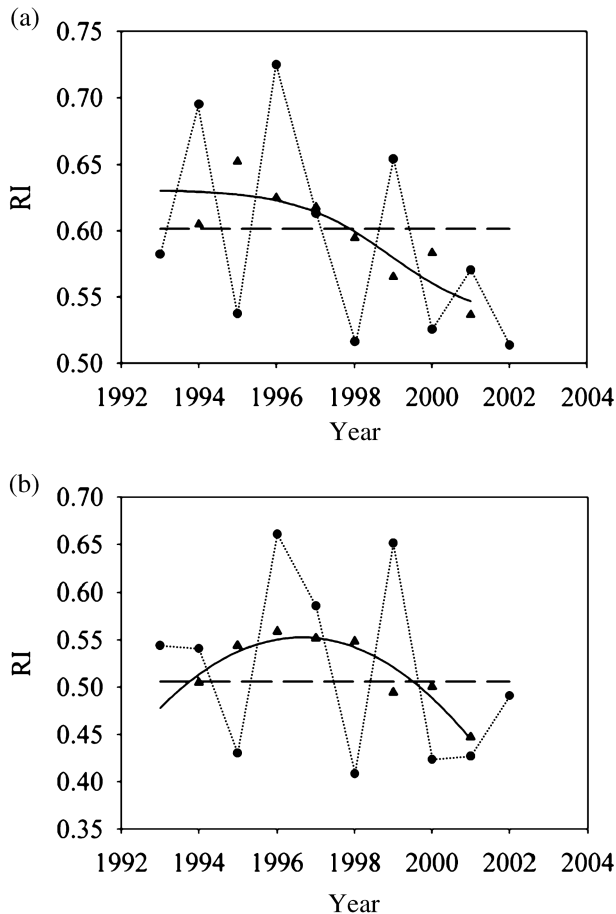


Fig. 4 Similarity between the genetic clustering in 2002 and the demographic clusterings for different years as measured by the Rand index. (a) Using the best-matching demographic clustering for each year, and (b) using the same number of demographic clusters as there were genetic clusters (see further explanation in the text). The yearly values are shown by dots connected by a dotted line. A 3-point moving average is shown by triangles, to which polynomial curves were fitted (continuous lines). The 3-point moving average is the average of three consecutive numbers with the year indicated as the mid-point, and it is used to smooth out short-term fluctuations to highlight longer term trends. The dashed lines show the similarity between the genetic clustering in 2002 and the spatial configuration of the habitat, with no regard to the presence of local populations (see Materials and methods).

clustering for 2002 and the demographic clusterings in 1995–1996, when the correspondence between the genetic and demographic clusterings peaked (Fig. 4).

Comparison between the demographic and genetic spatial clusterings in Fig. 4 suggests that the present genetic structure bears a signature of past demographic processes. Assuming such a time lag, we could expect that genetic diversity among the genetic clusters is also affected by past demographic patterns. To test this, we regressed the average

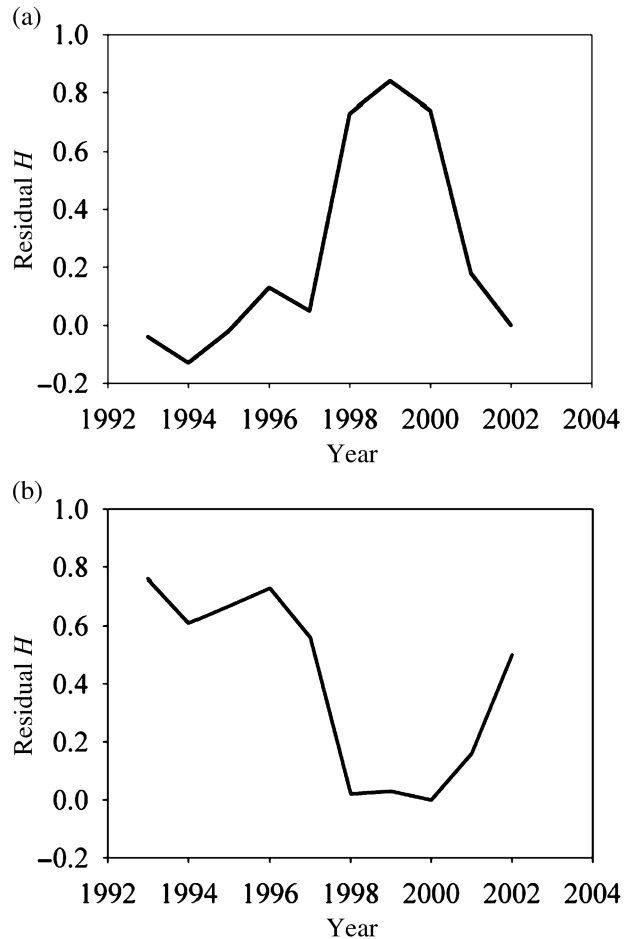


Fig. 5 Correlation between residual heterozygosity (H) and metapopulation size (N) in the years from 1993 to 2002, where residual heterozygosity is (a) the residual from the regression between H and N_{2002} or (b) the residual from the regression between H and N_{2000} . See the text for further explanation.

observed heterozygosity (H_0) calculated for microsatellite markers against the total number of larval families in the corresponding genetic cluster (N) in the years from 1993 until 2002. Forcing the current metapopulation size (N_{2002}) into the model, the best multiple regression included N_{2002} and N_{1999} , and explained 73% of variation in the H values (both explanatory variables were significant, $F = 10.4$, $P = 0.017$). The delayed effect is clearly seen in Fig. 5(a), which shows the correlations between N_{1993} and N_{2001} and the residual from the regression of H against N_{2002} . An even better model was obtained without forcing the current metapopulation size into the model, in which case N_{1993} and N_{2000} were selected and explained 79% of variation in the H values ($F = 13.9$, $P = 0.009$). The corresponding plot of the correlation coefficients between N_{1993} and N_{2002} and the residual from the regression of H against N_{2000} is shown in Fig. 5(b). The patterns in Fig. 5 were similar

when regression coefficients rather than correlation coefficients were examined. Naturally, these results can only be approximations, because different genetic clusters have partly unique demographic histories and may therefore exhibit unique temporal delays in the response of genetic diversity. Nonetheless, these results provide strong qualitative support for the presence of delayed effects that last for many years.

Discussion

Null alleles at microsatellite and SNP loci

Microsatellites are widely used in population genetic studies because they are highly variable, more likely to be neutral than many other markers, and the results are generally reproducible (Jarne & Lagoda 1996). However, sequencing studies indicate that variation occurs at non-negligible rates in the flanking regions of microsatellites (Angers & Bernatchez 1997; Grimaldi & Crouau-Roy 1997). Such variation may prevent primer annealing to template DNA during amplification, resulting in a null allele (Dakin & Avise 2004). Null alleles have been reported to be exceptionally frequent in some insect taxa such as Lepidoptera (reviewed in Meglecz *et al.* 2004), Diptera (Lehmann *et al.* 1997) and Orthoptera (Chapuis *et al.* 2005). Reasons may include lower stability of the flanking regions of microsatellites than of other genomic regions (Angers & Bernatchez 1997; Grimaldi & Crouau-Roy 1997; Meglecz *et al.* 2004). In most microsatellite studies on Lepidoptera, the presence of null alleles has been manifested as significant departures from the Hardy–Weinberg equilibrium due to heterozygote deficiency (Meglecz & Solignac 1998; Keyghobadi *et al.* 1999; Keyghobadi *et al.* 2005a). The Glanville fritillary is no exception (Palo *et al.* 1995; Sarhan 2006; present study).

The presence of null alleles may bias the analysis of genetic diversity. However, not all genetic measures are equally affected by null alleles (Dewoody *et al.* 2006; Wagner *et al.* 2006; Chapuis & Estoup 2007). In general, the *F*-statistics are noticeably affected (Dakin & Avise 2004), whereas Bayesian analysis of genetic spatial structure is less affected. *F*-statistics are especially sensitive because they make quantitative inferences about the genetic population structure by numerically quantifying the estimated extent of gene flow. The Bayesian model used in this study, and other comparable models, specifies only qualitatively the presence and nature of the spatial structure without quantitative inference about the associated gene flow.

To explore the general effects of deviations from the Hardy–Weinberg equilibrium, we conducted a range of simulations with increasing heterozygote deficiency (results not shown). The power to detect underlying spatial structure by the Bayesian cluster analysis as implemented in BAPS was reduced with increasing deviation from the Hardy–

Weinberg equilibrium, but the estimated genetic structure was conservative even in the presence of strong deviations from the equilibrium. Ideally, to minimize erroneous conclusions due to the presence of scoring errors and null alleles, models should use the adjusted allele frequency, which however, is not possible by the software commonly used to estimate *F*-statistics and isolation by distance (Dewoody *et al.* 2006). Simulation studies have shown that the bias introduced by null alleles is negligible when their frequency is less than 0.2 (Dakin & Avise 2004). In the present study, the frequency of null alleles ranged from 0.05 (CINXIA35) to 0.18 (CINXIA19), with only CINXIA1A having a frequency (0.29) greater than 0.2.

While the literature about null alleles affecting microsatellite loci is flourishing, there is hardly any literature about null alleles in SNPs, probably because of the still limited use of SNPs in ecology (Seddon *et al.* 2005). Six of our 10 SNPs showed a significant departure from the Hardy–Weinberg expectation. One locus (42.1) showed an excess of heterozygotes, while five loci showed heterozygote deficiency.

A BLAST search conducted in the Silkworm Knowledgebase (Silk DB, <http://silkworm.genomics.org.cn/index.jsp>) and National Center for Biotechnology Information (NCBI) public databases showed that the SNP loci used in this study are located in genomic regions with high variation. None of the loci matched any protein-coding genomic regions. This implies that while these regions are appropriate for population genetic studies, high nucleotide variation may affect the annealing of the primers to the DNA template, possibly yielding null alleles. We repeated the AMOVA after excluding loci with departure from the Hardy–Weinberg equilibrium, but the results remained essentially unchanged (data not shown), suggesting that they are not sensitive to the presence of null alleles.

Tests of the Hardy–Weinberg equilibrium performed for samples originating from separate genetic clusters with homogeneous allele frequencies (Table S1) yielded essentially the same results as tests using the pooled material. We therefore conclude that the deviations from the Hardy–Weinberg equilibrium are not likely to be due to spatial heterogeneity across the study area.

Spatial scale of genetic differentiation

For nearly 80 years, population geneticists have investigated genetic diversity within and between populations using Wright's *F*-statistics (Wright 1931), which are commonly used to draw inferences about gene flow and genetic drift (Kimura & Maruyama 1971; Slatkin 1985). The *F*-statistics are informative (though see Hedrick 1999) when local populations are relatively large and stable and inhabit discrete and equally connected habitats (Manel *et al.* 2003), but they are less helpful when individuals are distributed

more continuously in space, when there is fast turnover of small local populations as in classic metapopulations (Withlock & McCauley 1999), or when the metapopulation occupies a heterogeneous landscape that leads to a complex pattern of gene flow among dissimilar local populations.

We detected weak isolation by distance at the scale up to 35 km (data not shown). This result is comparable to the pattern reported by Saccheri *et al.* (2004, Fig. 10.3) for the same species using allozymes and microsatellites, although their result included a tail of larger values. The study of Saccheri *et al.* (2004) covered the entire Åland Islands, including populations from other islands apart from the main island, which explains the largest pairwise F_{ST} values. Moreover, the use of different genetic markers, allozymes vs. SNPs, may have contributed (Hedrick 1999) to the somewhat larger values in the previous than in the current study. Weak isolation by distance in the case of the Glanville fritillary is due to high gene flow among populations within networks (below) and less frequent gene flow between adjacent networks (Fig. 3).

The genetic cluster analysis identified groups of local populations with no significant differentiation in allele frequencies within groups (Fig. 3). Homogeneous genetic units thus identified were much greater than individual local populations, corresponding in spatial scale to the semi-independent patch networks used in previous ecological metapopulation studies of the Glanville fritillary (Hanski *et al.* 1995; Hanski 1999). These networks were defined as clusters of habitat patches in the landscape (Hanski *et al.* 1995). Network-scale homogeneous genetic units make good biological sense, because individual local populations are mostly so well connected within patch networks that there is substantial migration and gene flow. Using a dispersal model parameterized with empirical mark-recapture data, Hanski *et al.* (2000) estimated that more than half of the total butterfly days within a patch network are spent in a habitat patch other than the natal patch. The average lifetime movement distance is around 1.5 km (Ovaskainen & Hanski 2004), which is less than the average radius of the genetic clusters, but over a few generations there can be substantial gene flow across an entire patch network and genetic cluster.

Spatial scale of 'populations' in metapopulation biology

Given the relatively large homogeneous clusters of local populations, one might argue that the proper spatial units for population dynamics in this metapopulation are the networks or the genetic clusters rather than individual local populations. Nonetheless, although there is no significant genetic differentiation between nearby local populations, it is essential for ecological, genetic and evolutionary studies to take into account the fine-scale spatial structure at the level of local populations. In terms of demographic

dynamics, the density-dependent regulating processes operate at the level of local populations, including occasional food shortages (Hanski 1999), parasitism by specialist parasitoids (van Nouhuys & Hanski 2002), and habitat quality-dependent emigration and immigration (Kuussaari *et al.* 1996) potentially leading to an Allee effect (Kuussaari *et al.* 1998).

Considering genetic effects on population dynamics, the level of inbreeding is often so high in the mostly small local populations (Haikola *et al.* 2001) that the risk of local extinction becomes elevated by inbreeding depression (Saccheri *et al.* 1998). Our results on the division of genetic variance among larval groups, local populations and genetic clusters support the presence of substantial inbreeding (Table 3). High incidence of inbreeding may appear unlikely in the presence of extensive gene flow, but in reality many small populations, often consisting of just a single family group of full sibs, do not receive any immigrants in a given generation. Around 100 new and initially mostly very small populations are established every year (Hanski 1999), and many of them do not happen to receive any further immigration in the following year, thus leading to a generation of close inbreeding following population establishment.

Finally, there is a well-documented example of balancing selection at the phosphoglucose isomerase (*Pgi*) locus in this metapopulation (Haag *et al.* 2005; Hanski & Saccheri 2006), with opposing selection pressures at the within-population and metapopulation (patch network) levels. The evolutionary dynamics of *Pgi* would be entirely missed if pooled data for patch networks would be analysed instead of data for individual populations. In summary, although identifying genetically homogeneous spatial units can be helpful for many purposes, such as, for example, the identification of conservation units, such spatial units do not delimit 'local populations' for all purposes.

Genetic and demographic spatial structures

Our results demonstrate that the spatial genetic structure of the Glanville fritillary in a particular year was better explained by the past rather than by the concurrent demographic spatial structure or by the spatial configuration of the habitat in the landscape. Similarly, the observed heterozygosity within genetic clusters was to a large extent explained by the past size of the respective metapopulation. The natural explanation is that the genetic patterns are generated by demographic events, and the signature of major demographic events on genetic structure and diversity declines only slowly with time. In the present case, this 'time lag', as estimated by the best match between the demographic and genetic spatial clusterings, was 6–7 years, which is not very long but is consistent with the fast population turnover in this metapopulation (Hanski 1999; Nieminen *et al.* 2004). There were particularly great spatially

correlated changes in the metapopulation structure in 1997–1998, at which point populations in the northwestern part of the study area increased and populations in the southwest declined (Fig. 2). Such large-scale spatially correlated perturbations violate the assumptions of conventional interpretations of isolation-by-distance plots for neutral markers (Hutchinson & Templeton 1999), based on the forces of local genetic drift and gene flow, and are likely to lead to shallow isolation-by-distance relationships, as we observed in the present case.

Some previous studies have demonstrated that the correlation between genetic and geographical distances is improved by taking into account landscape structure (e.g. the moor frog (Vos *et al.* 2001), and the land snail (Arnaud 2003). Other studies have documented a discrepancy between demographic and genetic spatial patterns, mostly concerning genetic spatial structure of species in landscapes fragmented by human land use (Petit & Burel 1998; Van Dongen *et al.* 1998; Keyghobadi *et al.* 1999; Coulon *et al.* 2004; Holzhauer *et al.* 2006). In the case of invasive species, it is often the case that invasion history explains patterns in the genetic population structure (Grapputo *et al.* 2005; Holzer *et al.* 2006; Schmid-Hempel *et al.* 2007), especially at large spatial and temporal scales, for instance in the postglacial invasion of species into northern Europe (Solignac *et al.* 2005; Taylor & Keller 2007). To our knowledge there have been no previous attempts to examine the effects of both the spatial structure of the habitat and the demographic history in shaping the genetic spatial structure of metapopulations at the landscape scale.

The better match between the genetic spatial structure and the past rather than the present demographic spatial structure is indicative of large-scale demographic events in the past. In the absence of large-scale (network-level) perturbations, the metapopulation would be expected to remain close to an extinction–colonization equilibrium, and the network-level genetic structure should also be at equilibrium. Large-scale perturbations, due to spatially correlated changes in metapopulation size, are well documented for the Glanville fritillary metapopulation (Hanski 1999; Nieminen *et al.* 2004; Hanski & Meyke 2005) and are also evident in Fig. 2 in the present study. Most empirical studies of metapopulations and of habitat fragmentation (Keyghobadi *et al.* 1999; Harper *et al.* 2003; Coulon *et al.* 2004; Johansson *et al.* 2006; Zamudio & Wiczorek 2006) have failed to address such large-scale spatial dynamics. Large-scale perturbations are likely to be rare events in most systems, and it is unlikely that any general results concerning their consequences could be reached beyond the qualitative conclusion about mismatch between concurrent genetic and demographic spatial structures. The take-home message nonetheless is that genetic spatial structures in classic metapopulations may typically be in a transient state.

Acknowledgements

We thank J.M. Cano Arias, Gabriele Procaccini, Marjo Saastamoinen, Celine Teplitsky and three anonymous referees for comments on the manuscript, and Toshka Nyman and Evgeniy Meyke for technical assistance. This study has been funded by the Academy of Finland (grant numbers 38604 and 44887, Finnish Centre of Excellence Programme 2000–2005).

References

- Angers B, Bernatchez L (1997) Complex evolution of salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution*, **14**, 230–238.
- Arnaud JF (2003) Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. *Landscape Ecology*, **18**, 333–346.
- Batley J, Hayes PK (2003) Development of high throughput single nucleotide polymorphism genotyping for the analysis of *Nodularia* (Cyanobacteria) population genetics. *Journal of Phycology*, **39**, 248–252.
- Burel F, Baudry J, Butet A *et al.* (1998) Comparative biodiversity along a gradient of agricultural landscape. *Acta Oecologica*, **19**, 47–60.
- Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.
- Chapuis MP, Loiseau A, Michalakis Y, Lecoq M, Estoup A (2005) Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta Migratoria*. *Molecular Ecology Notes*, **5**, 554–557.
- Corander J, Marttinen P (2006) Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology*, **15**, 2833–2843.
- Corander J, Sirén J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, **23**, 111–129.
- Coulon A, Cosson JF, Angibault JM *et al.* (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology*, **13**, 2841–2850.
- Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504–509.
- Davies N, Villablanca FX, Roderick GK (1999) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology & Evolution*, **14**, 17–21.
- Deussen O, Hiller S, van Overveld C, Strothotte T (2000) Floating points: a method for computing stipple drawings. Computer Graphics Forum. *Proceedings of Eurographics*, **19**, 41–51.
- Dewoody JD, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellites data from wild populations. *Molecular Ecology Notes*, **6**, 951–957.
- Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.
- Ehrlich PR, Hanski I (2004) *On the Wings of the Checkerspot: A Model System for Population Biology*. Oxford University Press, New York.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

- Grapputo A, Boman S, Lindstrom L, Lyytinen A, Mappes J (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. *Molecular Ecology*, **14**, 4207–4219.
- Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. *Journal of Molecular Evolution*, **44**, 336–340.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.
- Guo S, Thompson E (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.
- Haag CR, Saastamoinen M, Marden J, Hanski I (2005) A candidate locus for variation in dispersal rate in a butterfly metapopulation. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2449–2456.
- Haikola S, Fortelius W, O'Hara RB *et al.* (2001) Inbreeding depression and the maintenance of genetic load in *Melita cinxia* metapopulations. *Conservation Genetics*, **2**, 325–335.
- Hanski I (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology*, **63**, 151–162.
- Hanski I (1999) *Metapopulation Ecology*. Oxford University Press, New York.
- Hanski I, Gaggiotti OE (2004) *Ecology, Genetics, and Evolution of Metapopulations*. Elsevier Academic Press, Amsterdam, The Netherlands.
- Hanski I, Meyke E (2005) Large-scale dynamics of the Glanville fritillary butterfly: landscape structure, population processes, and weather. *Annales Zoologici Fennici*, **42**, 379–395.
- Hanski I, Saccheri I (2006) Molecular-level variation affects population growth in a butterfly metapopulation. *PLoS Biology*, **4**, e129.
- Hanski I, Pakkala T, Kuussaari M, Lei G (1995) Metapopulation persistence of an endangered butterfly in a fragmented landscape. *Oikos*, **72**, 21–28.
- Hanski I, Moilanen A, Pakkala T, Kuussaari M (1996) The quantitative incidence function model and persistence of an endangered butterfly metapopulation. *Conservation Biology*, **10**, 578–590.
- Hanski I, Alho J, Moilanen A (2000) Estimating the parameter of survival and migration of individuals in metapopulations. *Ecology*, **81**, 239–251.
- Harper GL, Maclean N, Goulson D (2003) Microsatellite markers to assess the influence of population size, isolation and demographic change on the genetic structure of the UK butterfly *Polyommatus bellargus*. *Molecular Ecology*, **12**, 3349–3357.
- Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.
- Holzer B, Chapuisat M, Kremer N, Finet C, Keller L (2006) Unicoloniality, recognition and genetic differentiation in a native Formica ant. *Journal of Evolutionary Biology*, **19**, 2031–2039.
- Holzhauser SIJ, Ekschmitt K, Sander A-C, Dauber J, Wolters V (2006) Effect of historic landscape change on the genetic structure of the bush-cricket *Metrioptera roeseli*. *Landscape Ecology*, **21**, 891–899.
- Hubert L, Arabie P (1985) Comparing partitions. *Journal of Classification*, **2**, 193–218.
- Hutchinson WF, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.
- Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, **11**, 424–429.
- Johansson M, Primmer CR, Merila J (2006) History vs. current demography: explaining the genetic population structure of the common frog (*Rana temporaria*). *Molecular Ecology*, **15**, 975–983.
- Keyghobadi N, Roland J, Strobeck C (1999) Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology*, **8**, 1481–1495.
- Keyghobadi N, Roland J, Matter SF, Strobeck C (2005a) Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 553–560.
- Keyghobadi N, Roland J, Strobeck C (2005b) Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology*, **14**, 1897–1909.
- Kimura M, Maruyama T (1971) Pattern of neutral polymorphism in a geographically structured population. *Genetics Research*, **18**, 125–131.
- Kuussaari M, Nieminen M, Hanski I (1996) An experimental study of migration in the Glanville fritillary butterfly *Melitaea cinxia*. *Journal of Animal Ecology*, **65**, 791–801.
- Kuussaari M, Saccheri I, Camara M, Hanski I (1998) Allee effect and population dynamics in the Glanville fritillary butterfly. *Oikos*, **82**, 384–392.
- Lehmann T, Besansky NJ, Hawley WA *et al.* (1997) Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology*, **6**, 243–253.
- Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, **15**, 237–240.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
- Meglecz E, Solignac M (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas*, **128**, 179–180.
- Meglecz E, Petenian F, Danchin E *et al.* (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology*, **13**, 1693–1700.
- Neve G, Meglecz E (2000) Microsatellite frequencies in different taxa. *Trends in Ecology & Evolution*, **15**, 376–377.
- Nieminen M, Siljander M, Hanski I (2004) Structure and dynamics of *Melitaea cinxia* metapopulations. In: *On the Wings of the Checkerspot: A Model System for Population Biology* (eds Ehrlich PR, Hanski I), pp. 63–91. Oxford University Press, New York.
- van Nouhuys S, Hanski I (2002) Colonization rates and distances of a host butterfly and two specific parasitoids in a fragmented landscape. *Journal of Animal Ecology*, **71**, 639–650.
- van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.
- Orsini L, Pajunen M, Hanski I, Savilahti H (2007) SNP discovery by mismatch-targeting of Mu transposition. *Nucleic Acid Research*, **35**, e44.
- Ovaskainen O, Hanski I (2003) How much does an individual habitat fragment contribute to metapopulation dynamics and persistence? *Theoretical Population Biology*, **64**, 481–495.

- Ovaskainen O, Hanski I (2004) From individual behavior to metapopulation dynamics: unifying the patchy population and classic metapopulation models. *American Nature*, **164**, 364–377.
- Palo J, Vario S-L, Hanski I, Väinölä R (1995) Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. *Heredity*, **123**, 295–300.
- Petit S, Burel F (1998) Effect of landscape dynamics on the metapopulation of a ground beetle (Coleoptera, Carabidae) in a hedgerow network. *Agriculture Ecosystems and Environment*, **69**, 243–252.
- Ray C (2001) Maintaining genetic diversity despite local extinctions: a spatial scaling problem. *Biological Conservation*, **100**, 3–14.
- Raymond M, Rousset F (1995) GENEPOP version 1.2: a population genetic software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Saastamoinen M (2007) Life-history, genotypic, and environmental correlates of clutch size in the Glanville fritillary butterfly. *Ecological Entomology*, **32**, 235–242.
- Saccheri IJ, Kuussaari M, Kankare M *et al.* (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature*, **392**, 491–494.
- Saccheri IJ, Boggs CL, Hanski I, Ehrlich PR (2004) Genetics of checkerspot populations. In: *On the Wings of the Checkerspots* (eds Ehrlich PR, Hanski I), pp. 199–218. Oxford University Press, New York.
- Sarhan A (2006) Isolation and characterization of five microsatellite loci in the Glanville fritillary butterfly. *Molecular Ecology Notes*, **6**, 163–164.
- Schmid-Hempel P, Schmid-Hempel R, Brunner PC, Seeman OD, Allen GR (2007) Invasion success of the bumblebee, *Bombus terrestris*, despite a drastic genetic bottleneck. *Heredity*, **99**, 414–422.
- Seber GAF (1984) *Multivariate Observations*, Wiley & Sons, New York.
- Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology*, **14**, 503–511.
- Slatkin M (1985) Gene flow in natural populations. *Annual Review Ecology and Systematics*, **16**, 393–430.
- Solignac M, Cornuet JM, Vautrin D *et al.* (2005) The invasive Korea and Japan types of *Varroa destructor*, ectoparasitic mites of the Western honeybee (*Apis mellifera*), are two partly isolated clones. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 411–419.
- Solokov BP (1990) Primer extension technique for the detection of single nucleotides in genomic DNA. *Nucleic Acid Research*, **18**, 3671.
- Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, **61**, 334–345.
- Van Dongen S, Backeljau T, Matthysen E, Dhondt AA (1998) Genetic population structure of the winter moth (*Operophtera brumata* L.) (Lepidoptera, Geometridae) in a fragmented landscape. *Heredity*, **80**, 92–100.
- Vos CC, Antonisse-De Long AG, Goedhart PW, Smulders MJM (2001) Genetic similarity as a measure of connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity*, **86**, 598–608.
- Wagner AP, Creel S, Kalinowski ST (2006) Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity*, **97**, 336–345.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Withlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST} not equal $1/(4Nm + 1)$. *Heredity*, **82**, 117–125.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **10**, 97–159.
- Zamudio KR, Wieczorek AM (2006) Fine-scale genetic structure and dispersal among spotted salamander (*Ambystoma maculatum*) breeding populations. *Molecular Ecology*, **16**, 257–274.

Luisa Orsini is a postdoctoral research fellow in the Metapopulation Research Group at the University of Helsinki. Her research interests include molecular ecology, molecular evolution, speciation, and empirical population genetics. Jukka Corander is a professor of statistics at Åbo Akademi University. His research interests include general Bayesian theory of statistical inference and statistical methods for bioinformatics. Aino Alasentie worked as a research assistant during the present project. Ilkka Hanski is the director of the Metapopulation Research Group and research professor at the Department of Biological and Environmental Sciences in the University of Helsinki. His research interests include metapopulation biology and conservation biology.

Supplementary material

The following supplementary material is available for this article:

Table S1 Observed (H_O) and expected (H_E) heterozygosities, and the significance of the deviation from the Hardy-Weinberg equilibrium calculated separately for genetic clusters as identified by the program BAPS. Bold face indicates significant departure from the Hardy-Weinberg equilibrium.

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03782.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.